# Professionals underestimate patients' pain: comprehensive review

Tim Seers[1, 2], Sheena Derry[2], Kate Seers[3], R Andrew Moore[2]

1. Faculty of Medicine, Imperial College London, South Kensington, London, SW7 2AZ, UK (tim.seers@doctors.org.uk)
2. Pain Research and Nuffield Division of Anaesthetics, University of Oxford, The Churchill, Oxford, OX3 7LE, UK (sheena.derry@ndcn.ox.ac.uk, andrew.moore@ndcn.ox.ac.uk)
3. Warwick Research in Nursing, Warwick Medical School, University of Warwick, CV4 7AL, UK (Kate.Seers@warwick.ac.uk)

Corresponding author:
Dr RA Moore
Pain Research and Nuffield Division of Anaesthetics, Nuffield Department of Clinical Neurology, University of Oxford, The Churchill, Oxford OX3 7LE

Tel: +44 1865 225674
Email: andrew.moore@ndcn.ox.ac.uk

Number of text pages: 40
Number of Figures: 0
Number of Tables: 3
Supplementary files: 3

**Abstract**

Pain assessment by patient is the rule in clinical trials, but may not be in clinical practice. We examined studies comparing patient and professional assessment of pain in clinical practice using published studies (1990-2016; ≥20 patients), in English, comparing pain assessment within 24 hours by patient and healthcare professional. A difference of at least 10% of the maximum score was considered significant. We judged quality on sampling method, blinding, and study size.

Eighty studies (20,496 patients) provided data from a range of settings and locations; most (51%) used unbiased sampling, and most (68%) were blind or probably blind. Nine studies with ≥500 patients involved 58% of patients; 60 with <200 patients involved 25%. Large studies were more likely to use comprehensive or random sampling, and blinding of patient and professional.

Underestimation of pain by professional compared to patient was reported by 62/80 studies (78%), while there was no difference in 17 (21%), and overestimation in one (1%). Underestimation was reported in 75% of large studies (>500 patients), 91% of mid-sized studies (200-400), and 78% of small studies (<200). High quality studies (blind, comprehensive or random sampling, >200 patients) consistently reported underestimation (10/11; 91%). The extent of underestimation tended to increase with pain severity.

Professionals consistently tend to underestimate pain as compared to assessment by the patient. This tendency is more pronounced with more severe pain, and the extent of underestimation can be large. It is likely that this contributes to under-treatment of pain.

**Keywords:**

pain, measurement, professional, patient, concordance

## Introduction

The development of modern clinical analgesic research was characterised by intense scrutiny of methodology to make results reliable and free from bias. Key elements identified early were randomisation, blinding, and a minimum pain level. Because of the subjective nature of pain, another key element was that patients made their own assessment of pain intensity or relief [3,4,22,28,38]. No objective test of pain has been developed, and measurement by the patient has been the rule in clinical trials to this day; early methodological criteria have stood the test of time [35].

Why then do journals carry substantial numbers of papers (around 20 in the past 10 years) examining the question of differences in pain assessment between patient and professional? Such studies may reflect some aspect of clinical practice such as a test of new record systems [18], quality improvement systems [61], or a clinical classification [6]. Anecdotally, professional rather than patient assessment of pain can appear to be the norm in everyday practice.

Professionals can have very different views from patients about what is important. In rheumatoid arthritis, for example, patients consider that pain is the most important symptom but rheumatologists consider the number of swollen joints to have primacy [57], leading to significant discordance in patient and professional global assessments [13]. In advanced cancer, agreement between patient and observer ratings for various measures, including pain, was generally modest to low [55]. Patients are often not consulted over their pain, even when pain is to be expected, after surgery for example. A third of respondents to the large European PATHOS study said that pain was not assessed, and 60% had no threshold pain score above which analgesics should be given [5], despite low pain being a driver for patient satisfaction [37]. Chronic pain is more complicated, but low pain is associated

with better economic outcomes, including work [39,45]. Even in standardised settings, pain comes at different times, at different intensities, and sometimes not at all [34]. Those experiencing moderate or severe pain consistently desire pain reduced to no pain or only mild pain, and quickly [42]. Members of the general public equate moderate, and especially severe, pain with large decrements in health utility values, irrespective of their own pain status [14].

It is over 15 years since a narrative review of the agreement between patient and professional pain ratings concluded: "*health professionals are not particularly accurate in their pain assessments and often tend to underestimate pain in others*" [53]. Pain is underestimated [25]. Why this is so is unclear. Imaging studies suggest that the human brain can detect pain intensity of others from facial expressions [51]; empathy for pain may be underpinned by neural structures involved in the experience of pain [27]. Exhaustive review of factors affecting social judgement of pain describes its complexity [58].

In order to obtain a more complete assessment of concordance in pain rating between patients and professionals, we conducted a systematic review across all pain states, in adults able to communicate.

## Methods

### Searching

We searched electronic databases (MEDLINE, EMBASE, CINAHL and PsychInfo) using the terms: 'Pain measurement OR (pain measurement).mp AND (agreement OR disagreement OR concordance OR discordance OR judge* OR underestimat* OR overestimat*). Searching was limited to the period 1990 to May 2016 to ensure the contemporary value of comparisons or conclusions. Electronic searches were run independently by two authors.

Because observational studies are difficult to identify electronically we supplemented electronic searches with extensive eclectic searching using reference lists, review articles, and suggestive applications in PubMed and Google Scholar [31,50]. For example, titles of studies considered possible includes were pasted into PubMed, and the lists of 'Similar articles' and 'Cited by PubMed articles' were examined to see if any other articles fulfilled the inclusion criteria. The same process was followed with Google Scholar, in that case examining the citation list for the index article. For any probable include found by this method, a similar process was followed. Extensive eclectic searching was performed by all authors in different ways.
All likely articles were the discussed by all authors to choose an agreed list of included articles.

We included only English language studies because we did not have expertise to translate nuanced information. We did not consider short abstracts or unpublished studies, or contact authors.

### Criteria for inclusion

Included studies had to fulfil the following criteria:

i. Used valid pain assessment method (categorical verbal rating scale [VRS], numerical rating scale [NRS], or visual analogue scale [VAS]).

ii. Conducted in an institutional setting, comparing pain assessment between a patient and any recognised healthcare professional, with paired assessments made within 24 hours.

iii. Patient able to understand and convey their rating of pain in the context of an appropriate pain assessment tool (for instance,

scoring ≥23 on the mini-mental state examination (MMSE), or not having severe trauma or dementia.

iv. Study population was ≥18 years of age.

v. Study involved ≥20 patients, because small size is highly affected by the random play of chance [40,43,59].

Two authors independently reviewed the titles and abstracts to determine whether they satisfied the inclusion criteria; disagreements were resolved through discussion between all authors. Review authors were not blinded at any stage to author's names and institutions, journal of publication, or study results.

### Assessment of methodological quality

We made no formal assessment of the risk of bias in studies. We anticipated that we would find mainly small observational studies that were heterogeneous in population, setting, and methods. Deeks et al considered that biases in non-randomised studies are highly variable in causation, direction, and magnitude [12]. They are thought of as introducing additional uncertainty, rather than an estimable systematic bias [21].

Our interest was in the assessment of pain intensity by two or more individuals, not an intervention to change a patient's condition. We therefore considered the most important criteria to evaluate were:

1. Study recruitment. We considered that studies encompassing a whole population were less likely to be subject to any selection bias. Higher quality might then be expected from comprehensive sample or randomly selected populations than from a convenience population where selection might have occurred.

2. Blinding. We considered blinded studies where the two parties were unaware of the other's assessment to be of higher quality than unblinded studies.
3. Study size. Small studies are highly susceptible to random chance effects [40,43,59], and considerable biases have been observed in small studies [10,11,16,24,49,60]. We therefore looked for agreement or disagreement in studies of more than 500 patients (large), 200 to 500 patients (moderate), or fewer than 200 patients (small size, minimum 20).

### Outcome measures

We considered any attempt to provide an estimation of the level of disagreement between the patient self-reported pain and the pain recorded by the healthcare professional. This could be by stating the average discrepancy, reporting numbers within classes of discordance or concordance, or any appropriate statistical result. Where studies reported observations of multiple episodes of pain, we used the average pain intensity or highest reported pain intensity (or the closest possible approximation to this). Our preferred measure of disagreement was for patient-professional scores to differ by at least 10% of the pain scale being used; that would be 1 cm or 10 mm on a VAS, 1 point on a 0-10 point NRS [23], or the equivalent on any categorical or verbal rating scale. This 'Iafrati criterion' is widely reported in studies, though has been criticised as being somewhat arbitrary [47]. We also looked for any other measure of agreement or disagreement between patient and health professional. For each study all authors agreed a judgement on whether it reported evidence of underestimation, overestimation, or no difference between patient and professional.

### Data extraction and analysis

Data extracted from studies was entered into a standard form by one of two authors. This was then checked by another of the two authors, and where necessary, by a third review author. Any disagreements were resolved through discussion.

No statistical analysis was planned a priori. A descriptive analysis was considered likely to be the best approach because we expected considerable heterogeneity between studies in terms of pain condition, professionals involved, and pain assessment tools used, as well as the almost 25-year period over which they were conducted and published, together with the considerable variation in size. We used the GRADE approach to assess the quality of evidence [20].

## Results

### Description and quality of studies

Individual author searches tended to identify the same potentially relevant studies/articles. The inclusion criteria were met by 80 studies, with 20,496 patients provided relevant data (Table 1).

Most patients (58%) were in the nine studies involving over 500 patients, while 60 studies reporting on fewer than 200 patients involved only 25% of the total. A range of different pain states was included, with acute pain, cancer pain, chronic non-cancer pain, and pain in the community or nursing homes all represented. Settings for studies were diverse, involving whole hospitals or departments, or were limited to specific situations such as units for older people, critical care, burns, or nursing homes. Studies were predominantly set in Europe, the USA, Australia, or Canada (71/80 studies), but some were conducted in, Israel (3), Turkey (2), Iran (1), Kuwait (1), South Korea (1), and South Africa (1). References for the included studies are in

Supplementary file 1, and a detailed table describing the studies in terms of setting, location, sampling, and blinding in Supplementary file 2, together with a summary of the results in individual studies.

There was a tendency for older studies to be smaller (Supplementary file 3). The decade between 1990 and 1999 had 43% of all studies, but 24% of all patients. The decade between 2000 and 2009 had 41% of studies and 49% of patients. The period from 2010 to 2016 had 16% of studies but 27% of patients. Sixteen of the 20 studies involving more than 200 patients were published since 2000, and were conducted principally in the USA and Europe.

About half of studies used comprehensive (or probably comprehensive) sampling (41%), random sampling (4%) or were analyses of randomised trials (6%); the remainder were convenience samples (38%), or there was insufficient information to make a judgement (10%). Only one of the 20 studies involving more than 200 patients used a convenience sample compared with 29/60 studies involving fewer than 200.

Most studies (68%) were blind or probably blind, meaning that neither patient nor professional was aware of each other's pain assessment score; the largest studies (>500 patients) were more likely to be blind (78%). Only three studies (4%) were explicitly not blind, and in the remainder (18%) the method of blinding was unclear or not given (Table 1).

Pain assessments were usually made using standard VAS or NRS (100 mm or 10 cm) or VRS (4-point), although a few studies used non-standard variations, such as a 15 cm VAS, 3- or 6-point VRS, and a 21-point box scale. Some studies used more than one scale, and others converted one scale into another, for example, a VAS score into a VRS. A few studies reported only the presence or absence of pain, without any attempt to measure its intensity.

Almost all of the studies involved patients with a range of pain scores, from mild pain to severe pain.

Agreement or disagreement was formally assessed in 59 studies, most frequently in larger studies; definitions of discordance used in the studies are shown in Table 2. All definitions of disagreement were equivalent to a difference of at least 1 in 10, or 10% of the available score, and frequently higher, especially in the largest studies. While the level of disagreement was not assessed in the remaining 21 studies, information was provided that allowed a comparison to be made on the level of discrepancy between patient and professional, such as mean pain scores, or statistical comparisons.

### Degree of patient and professional agreement

Using these criteria, underestimation of pain by professional compared to the patient was reported by 62 of the 80 studies (78%; Table 3); 17 showed no difference, and in only a single study in burns patients was there any consistent overestimation [15].

Nine studies assessed between 563 and 3575 patients each (mean study size 1325, median 869). In seven, disagreement between patient and professional occurred in 10% to 68% of dyads, and group means disagreed by 10/100 to 30/100. Professional underestimation of pain compared to the patient was reported in six studies, ranging from "minor" (undefined) to nine times more frequent underestimation than overestimation. No clear direction was apparent in three studies, one of which reported overestimation for incident pain and underestimation for neuropathic pain. The four studies that were both blinded and had comprehensive or random selection reported significant underestimation. More frequent underestimation than overestimation was reported for higher levels of pain intensity in three studies that specifically looked for an effect [30,32,36].

Eleven studies assessed 200 to 465 patients each (mean study size 308, median 252). In six studies disagreement between patient and professional occurred in 33% to 68% of dyads but another reported 65% to 80% for patients with moderate or severe pain. Another reported a mean difference between patient and professional pain scores of 29/100. Professional underestimation of pain compared to the patient was reported in 10 studies, ranging from 1.5 to three times greater underestimation than overestimation or "significant" or "mainly underestimation"; in one study there was no clear direction. More frequent underestimation than overestimation was reported with higher pain intensity in one study. Six out of seven studies that were both blinded and had comprehensive or random selection reported significant underestimation.

Sixty studies assessed 20 to 198 patients each (mean study size 87, median 71). In 30 studies disagreement between patient and professional occurred in 15% to 98% of dyads. Professional underestimation of pain compared to the patient was reported in 46 studies, ranging from 1.5 to 10 times greater underestimation than overestimation, or from "minor" to "major" underestimation. No direction was reported in 13 studies. One study in burns patients reported overestimation. More frequent underestimation than overestimation was reported with higher pain intensity in 13 studies. Twenty-three out of 27 studies that were both blinded and had comprehensive or random selection reported significant underestimation.

### GRADE evaluation

Because of the observational design of the studies, our initial evaluation of the evidence was low quality. We uprated our assessment of quality to moderate-to-high for four reasons:

1. The magnitude of the degree of underestimation was large and consistent.
2. We could identify no confounding that would suggest a spurious cause for under- or overestimation.
3. There was a greater effect in studies of higher quality, irrespective of size. A greater proportion of studies reported underestimation where studies were both blind and had comprehensive or random sampling (33/38; 87% compared with 78% overall), and with 10/11 (91%) of studies with >200 patients reporting underestimation.
4. We judged there to be the equivalent of a dose-response in underestimation. A number of large studies examining a link between the degree of underestimation and the patient-reported pain severity showed greater underestimation in patients with more severe pain.

## Discussion

This systematic review has several strengths. Studies involved different countries and a broad range of clinical conditions where pain is frequent, such as hospital settings, outpatients, primary care, and nursing homes. There was a clear and consistent message, that professionals frequently underestimate pain compared with the patient report. Underestimation was reported in 75% of large studies, 91% of mid-sized studies, and 78% of small studies. Underestimation was more frequent in larger of higher quality studies. Findings were consistent in different clinical settings, and despite the use of different pain scales and varying definitions of agreement or disagreement. Consistent overestimation of pain by professionals was reported in a single small study. While there may be concerns over quality and size of the studies and the measurement and magnitude of the underestimation, the fact of generalised underestimation appears to be incontrovertible. This was

especially the case with higher levels of pain intensity. Our GRADE assessment indicated the evidence to be moderate-to-high quality.

Assessment of sources of bias is largely undefined for the type of studies in this review. We judged the most critical issues were sampling method, blinding of patient and professional to the other's score, and study size. Most studies (51%) used comprehensive (or probably comprehensive) sampling, random sampling, or analyses of randomised trials; 19/20 of the largest studies used unbiased sampling. Most studies (68%) were blind or probably blind, with only 4% clearly not blind. Few studies were large, with only 9 studies (but 58% of all patients) involving over 500 patients, and 74% of patients were in the 20 studies with over 200 patients. Large studies were more likely to use comprehensive or random sampling, and blinding of patient and professional. Large studies of the highest quality were more consistent in reporting underestimation.

A particular issue was the degree of difference required for a result to be judged a disagreement. All of the formal measures used a criterion of 10% of the value of the pain scale used (1/10, or 10/100). This is not a small difference. Farrar et al suggested that minimal clinically important differences in pain were around 33%, but that was an absolute change in an individual [17]. Differences of 10% or above in mean pain scores between active drug and placebo are indicative of 'effective' drugs because while some get good pain relief, some do not, even with the best drugs [41] Nevertheless, varying definitions of what constituted disagreement represents a weakness, though no reasonable explanation of changing definitions would affect the direction of the result. An analysis according to whether or not pain was at the acceptable level of no pain or no worse than mild pain is unlikely to have changed the findings, and may have made the comparison more stark [42].

A further weakness is the difficulty of ascertaining whether all appropriate studies have been identified. Searching for observational studies electronically has weaknesses, but we supplemented this with other methods found to be successful in similar circumstances [31,50]. We did not have the skills adequately to translate studies in languages other than English. An updated search in PubMed to May 2017 identified one additional study that satisfied our inclusion criteria. This study supports the general findings of this review, and in particular the greater underestimation of severe pain [52].

There is a clear message that health care professionals are often poor at assessing the pain of patients for whom they care. Our findings extend those of Kappesser and Williams, who reviewed 13 studies that looked at patient-observer agreement and found that professionals were more likely to underestimate patients' pain than were the patients' relatives [26]. A systematic review of 12 studies in rheumatoid arthritis (11,879 patients) reported considerable discordance between patient and professional rating of global assessment, with patient estimate of a worse state, driven by pain and function [13]. Children reported higher acute pain scores than their parents [33]. It has been argued that professionals can become "numb" to suffering [47]. Pain professionals may be better than other healthcare professionals in assessing patients' pain, at least in one example of paediatric intensive care nurses [29].

The reason why pain is underestimated is interesting. Patient and clinician use information largely inaccessible to the other: the patient uses his or her own beliefs about the pain and underlying causes; the clinician uses the patient's behaviour, facial expression, and information about the disorder presumed to cause the pain, and knowledge of the condition [58]. Perhaps different estimations are to be expected and may be resistant to change. Even when exposed to the patient's rating of pain after estimating it blind to that

rating, clinicians do not adjust their estimates much towards the patient's rating [48].

Understanding of causation is likely to help in finding ways to overcome underestimation by professionals [58]. Presently, however, the ***fact*** of significant underestimation is of primary importance. Pain assessment and report by the patient has been accepted as an absolute requirement for pain trials for over half a century [3,4,22,28,38]. A recent guideline on management of postoperative pain states "because pain is inherently subjective, patient self-report is the primary basis of all pain assessments" [7]. Moreover, studies consistently demonstrate that improvements in pain therapy in hospital and especially in postoperative patients come from pain assessment *by the patient*, action when pain breaches a threshold, and re-evaluation of pain *by the patient* [1,2,46,61]. Yet painful conditions are among the most common afflicting humans worldwide, and associated with considerable numbers of years living with disability [63], and surveys consistently show pain, including severe pain, to be a common experience even in hospital [19].

These studies were carried out between 1991 and 2016, with no indication that the situation is improving. There will be circumstances in which patients cannot communicate their pain, but where they can it is the patient's report that is important, as accepted for clinical trials for decades [22], and as current guidance makes clear in some institutions [54]. Where professionals do not ask or believe the patient, the result is that significant pain is undertreated in a large proportion of them [65]. The practical message is to have pain measured regularly by the patient where possible, to set thresholds for action, and to act on high pain scores, again as some guidance recommends [1,54].

Since professionals can only ever *estimate* the patient's pain, it may not be surprising that patient and professional actual pain scores do not always agree. Professionals should therefore avoid assuming that their estimated

score is accurate – the patient needs to be asked about their own pain. A concern is that healthcare professionals may value their own rating over that of the patient: they know the patient's rating but discount it.

## Conclusions and recommendations

1. Results of this systematic review emphasise that professionals cannot be sure of accurately assessing the pain experience of the patient before them. It is the patient's own experience that is important. In 1946 Beecher described eliciting pain levels from a wounded soldier by asking the question "As you lie there are you having any pain?" [4]. This would be a good first step.
2. Patient self-assessment of pain should be the rule in clinical practice, as it is in clinical trials.
3. Other than for quality improvement purposes [18,61], there is no compelling reason for continuing to perform and publish studies comparing pain scoring by patient and professional. With half a century of research consistently producing the same result is time to stop. Systematic research has identified similar circumstances previously [56].
4. New research might profitably take different directions to better understand the problem of pain underestimation by professionals. These may include, but are not restricted to investigating what impact it may have on patients and healthcare systems, its psychological underpinnings, the identification of its correlates, and perhaps pragmatic mechanisms to manipulate its magnitude or ameliorate its effects, for example along the lines of providing a pain-free hospital [62].
5. Pain scoring systems are tools to help. Categorical, visual analogue, and other scoring systems are generally well correlated [8,9].

6. Patients with pain consider a satisfactory result to be a pain level of no worse than mild pain, achieved quickly. Patient reported pain of moderate or severe (more than 30/100 mm on a 100 mm VAS) is usually a reason to act.

There is a pressing need for quality control to check that patients are asked about their pain levels and that what they themselves report is accurately recorded. That, together with examples of what can work in particular pain situations, especially in managing chronic non-cancer pain, is today's research agenda. We are abundantly aware of a problem, and in many circumstances we know the solution, but implementation is long overdue.

**Acknowledgements**

This work had no specific funding, other than with institutional support from the Oxford Pain Research Trust. All authors were involved with the initial discussions about the scope of the review and whether such a review was possible. TS and SD extracted data, which was checked and added to by KS and RAM. All authors were involved in the analyses and interpretation of available data, in the writing of the various drafts, and in approving the final version of the paper. No author has any known financial conflict of interest.

We are grateful for additional insights and advice provided by three anonymous peer reviewers.

## References


1. Aldington DJ, McQuay HJ, Moore RA. End-to-end military pain management. Philos Trans R Soc Lond B Biol Sci 2011;366:268-75.
2. Bardiau FM, Taviaux NF, Albert A, Boogaerts JG, Stadler M. An intervention study to enhance postoperative pain management. Anesth Analg 2003;96:179-85.
3. Beecher HK, Keats AS, Mosteller F, Lasagna L. The effectiveness of oral analgesics (morphine, codeine, acetylsalicylic acid) and the problem of placebo "reactors" and "non-reactors". J Pharmacol Exp Ther 1953;109:393-400.
4. Beecher HK. Pain in men wounded in battle. Ann Surg 1946;123:96-105.
5. Benhamou D, Berti M, Brodner G, De Andres J, Draisci G, Moreno-Azcoita M, Neugebauer EA, Schwenk W, Torres LM, Viel E. Postoperative Analgesic THerapy Observational Survey (PATHOS): a practice pattern study in 7 central/southern European countries. Pain 2008;136:134-41.
6. Brunelli C, Costantini M, Di Giulio P, Gallucci M, Fusco F, Miccinesi G, Paci E, Peruselli C, Morino P, Piazza M, Tamburini M, Toscani F. Quality-of-life evaluation: when do terminal cancer patients and health-care providers agree? J Pain Symptom Manage 1998;15:151-8.
7. Chou R, Gordon DB, de Leon-Casasola OA, Rosenberg JM, Bickler S, Brennan T, Carter T, Cassidy CL, Chittenden EH, Degenhardt E, Griffith S, Manworren R, McCarberg B, Montgomery R, Murphy J, Perkal MF, Suresh S, Sluka K, Strassels S, Thirlby R, Viscusi E, Walco GA, Warner L, Weisman SJ, Wu CL. Management of Postoperative Pain: A Clinical Practice Guideline From the American Pain Society, the American Society of Regional Anesthesia and Pain Medicine, and the American Society of Anesthesiologists' Committee on Regional Anesthesia, Executive Committee, and Administrative Council. J Pain 2016;17:131-57.

8. Collins SL, Edwards J, Moore RA, Smith LA, McQuay HJ. Seeking a simple measure of analgesia for mega-trials: is a single global assessment good enough? Pain. 2001;91:189-94.
9. De Conno F, Caraceni A, Gamba A, Mariani L, Abbattista A, Brunelli C, La Mura A, Ventafridda V. Pain measurement in cancer patients: a comparison of six methods. Pain 1994;57:161-6.
10. Dechartres A, Altman DG, Trinquart L, Boutron I, Ravaud P. Association between analytic strategy and estimates of treatment outcomes in meta-analyses. JAMA 2014;312:623-30.
11. Dechartres A, Trinquart L, Boutron I, Ravaud P. Influence of trial sample size on treatment effect estimates: meta-epidemiological study. BMJ 2013;346:f2304.
12. Deeks JJ, Dinnes J, D'Amico R, Sowden AJ, Sakarovitch C, Song F, Petticrew M, Altman DG; International Stroke Trial Collaborative Group; European Carotid Surgery Trial Collaborative Group. Evaluating non-randomised intervention studies. Health Technol Assess 2003;7(27):iii-x, 1-173.
13. Desthieux C, Hermet A, Granger B, Fautrel B, Gossec L. Patient-physician discordance in global assessment in rheumatoid arthritis: A systematic literature review with meta-analysis. Arthritis Care Res (Hoboken) 2016;68:1767-73.
14. Eldabe S, Lloyd A, Verdian L, Meguro M, Maclaine G, Dewilde S. Eliciting health state utilities from the general public for severe chronic pain. Eur J Health Econ 2010;11:323-30.
15. Everett JJ, Patterson DR, Marvin JA, Montgomery B, Ordonez N, Campbell K. Pain assessment from patients with burns and their nurses. J Burn Care Rehabil 1994;15:194-8.
16. Fanelli D, Costas R, Ioannidis JP. Meta-assessment of bias in science. Proc Nat Acad Sci USA 2017;114:3714-9.

17. Farrar JT, Portenoy RK, Berlin JA, Kinman JL, Strom BL. Defining the clinically important difference in pain outcome measures. Pain 2000;88:287-94.
18. Goulet JL, Brandt C, Crystal S, Fiellin DA, Gibert C, Gordon AJ, Kerns RD, Maisto S, Justice AC. Agreement between electronic medical record-based and self-administered pain numeric rating scale: clinical and research implications. Med Care 2013;51:245-50.
19. Gregory J, McGowan L. An examination of the prevalence of acute pain for hospitalised adult patients: a systematic review. J Clin Nurs 2016;25(5-6)583-98.
20. Guyatt GH, Oxman AD, Vist GE, Kunz R, et al; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924-6.
21. Higgins JP, Green S, editor(s). Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org. Chapter 13.5.1.2
22. Houde RW, Wallenstein SL, Rogers A. Clinical pharmacology of analgesics. 1. A method of assaying analgesic effect. Clin Pharmacol Ther 1960;1:163-74.
23. Iafrati NS. Pain on the burn unit: patient vs nurse perceptions. J Burn Care Rehabil 1986;7:413-6.
24. IntHout J, Ioannidis JP, Borm GF, Goeman JJ. Small studies are more heterogeneous than large ones: a meta-meta-analysis. J Clin Epidemiol 2015;68:860-9.
25. Kappesser J, Williams AC, Prkachin KM. Testing two accounts of pain underestimation. Pain 2006;124:109-16.
26. Kappesser J, Williams AC. Pain estimation: asking the right questions. Pain 2010;148:184-7.

27. Lamm C, Decety J, Singer T. Meta-analytic evidence for common and distinct neural networks associated with directly experienced pain and empathy for pain. Neuroimage 2011;54:2492-502.
28. Lasagna L, De Kornfeld TJ. Analgesic potency of normorphine in patients with postoperative pain. J Pharmacol Exp Ther 1958;124:260-3.
29. Latimer M, Jackson PL, Eugène F, MacLeod E, Hatfield T, Vachon-Presseau E, Michon PE, Prkachin KM. Empathy in paediatric intensive care nurses part 1: Behavioural and psychological correlates. J Adv Nurs 2017;73:2676-85.
30. Laugsand EA, Sprangers MA, Bjordal K, Skorpen F, Kaasa S, Klepstad P. Health care providers underestimate symptom intensities of cancer patients: a multicenter European study. Health Qual Life Outcomes 2010;8:104.
31. Lemeshow AR, Blum RE, Berlin JA, Stoto MA, Colditz GA. Searching one or two databases was insufficient for meta-analysis of observational studies. J Clin Epidemiol 2005;58:867-73.
32. Mäntyselkä P, Kumpusalo E, Ahonen R, Takala J. Patients' versus general practitioners' assessments of pain intensity in primary care patients with non-cancer pain. Br J Gen Pract 2001;51:995-7.
33. Matziou V, Vlachioti E, Megapanou E, Ntoumou A, Dionisakopoulou C, Dimitriou V, Tsoumakas K, Matziou T, Perdikaris P. Perceptions of children and their parents about the pain experienced during their hospitalization and its impact on parents' quality of life. Jpn J Clin Oncol 2016;46:862-70.
34. McQuay HJ, Bullingham RE, Moore RA, Evans PJ, Lloyd JW. Some patients don't need analgesics after surgery. J R Soc Med 1982;75:705-8.
35. McQuay HJ, Derry S, Eccleston C, Wiffen PJ, Moore RA. Evidence for analgesic effect in acute pain - 50 years on. Pain 2012;153:1364-7.
36. Melotti RM, Samolsky Dekel BG, Carosi F, Ricchi E, Chiari P, D'Andrea R, Di Nino G. Categories of congruence between inpatient self-reported pain and nurses evaluation. Eur J Pain 2009;13:992-1000.

37. Mhuircheartaigh RJ, Moore RA, McQuay HJ. Analysis of individual patient data from clinical trials: epidural morphine for postoperative pain. Br J Anaesth 2009;103:874-81.
38. Modell W, Houde RW. Factors influencing clinical evaluation of drugs; with special reference to the double-blind technique. J Am Med Assoc 1958;167:2190-9.
39. Moore RA, Derry S, Taylor RS, Straube S, Phillips CJ. The costs and consequences of adequately managed chronic non-cancer pain and chronic neuropathic pain. Pain Pract 2014;14:79-94.
40. Moore RA, Gavaghan D, Tramèr MR, Collins SL, McQuay HJ. Size is everything – large amounts of information are needed to overcome random effects in estimating direction and magnitude of treatment effects. Pain 1998;78:209-16.
41. Moore RA, Moore OA, Derry S, Peloso PM, Gammaitoni AR, Wang H. Responder analysis for pain relief and numbers needed to treat in a meta-analysis of etoricoxib osteoarthritis trials: bridging a gap between clinical trials and clinical practice. Ann Rheum Dis 2010;69:374-9.
42. Moore RA, Straube S, Aldington D. Pain measures and cut-offs – 'no worse than mild pain' as a simple, universal outcome. Anaesthesia 2013;68:400-12.
43. Nguyen TL, Collins GS, Lamy A, Devereaux PJ, Daurès JP, Landais P, Le Manach Y, Le Manach Y. Simple randomization did not protect against bias in smaller trials. Journal of Clinical Epidemiology 2017 Feb 29 [Epub ahead of print].
44. Nüesch E, Trelle S, Reichenbach S, Rutjes AW, Tschannen B, Altman DG, Egger M, Jüni P. Small study effects in meta-analyses of osteoarthritis trials: meta-epidemiological study. BMJ 2010;341:c3515.
45. Patel AS, Farquharson R, Carroll D, Moore A, Phillips CJ, Taylor RS, Barden J. The impact and burden of chronic pain in the workplace: a qualitative systematic review. Pain Pract 2012;12:578-89.

46. Pogatzki-Zahn E, Kutschar P, Nestler N, Osterbrink J. A prospective multicentre study to improve postoperative pain: identification of potentialities and problems. PLoS One 2015;10:e0143508.
47. Prkachin KM, Solomon PE, Ross J. Underestimation of pain by health-care providers: towards a model of the process of inferring pain in others. Can J Nurs Res 2007;39:88-106.
48. Riva P, Rusconi P, Montali L, Cherubini P. The influence of anchoring on pain judgment. J Pain Symptom Manage 2011;42:265-77.
49. Roberts I, Ker K, Edwards P, Beecher D, Manno D, Sydenham E. The knowledge system underpinning healthcare is not fit for purpose and must change. BMJ 2015;350:h2463.
50. Ruppen W, Derry S, McQuay H, Moore RA. Incidence of epidural hematoma, infection, and neurologic injury in obstetric patients with epidural analgesia/anesthesia. Anesthesiology 2006;105:394-9.
51. Saarela MV, Hlushchuk Y, Williams AC, Schürmann M, Kalso E, Hari R. The compassionate brain: humans detect intensity of pain from another's face. Cereb Cortex 2007;17:230-7.
52. Samolsky Dekel BG, Gori A, Vasarri A, Sorella MC, Di Nino G, Melotti RM. Medical Evidence Influence on Inpatients and Nurses Pain Ratings Agreement. Pain Res Manag 2016;2016:9267536.
53. Solomon P. Congruence between health professionals' and patients' pain ratings: a review of the literature. Scand J Caring Sci 2001;15:174-80.
54. Somerset J. Pain assessment and management guidance for all wards and units. Mid Essex Hospital Services NHS Trust Clinical Guideline 11027. September 2014.
55. Stone P, Gwilliam B, Keeley V, Todd C, Gittins M, Kelly L, Barclay S, Roberts C. Patients' reports or clinicians' assessments: which are better for prognosticating? BMJ Support Palliat Care 2012;2:219-23.
56. Straube S, Harden M, Schröder H, Arendacka B, Fan X, Moore RA, Friede T. Back schools for the treatment of chronic low back pain:

possibility of benefit but no convincing evidence after 47 years of research-systematic review and meta-analysis. Pain 2016;157:2160-72.
57. Studenic P, Radner H, Smolen JS, Aletaha D. Discrepancies between patients and physicians in their perceptions of rheumatoid arthritis disease activity. Arthritis Rheum 2012;64:2814-23.
58. Tait RC, Chibnall JT, Kalauokalani D. Provider judgments of patients in pain: seeking symptom certainty. Pain Med 2009;10:11-34.
59. Thorlund K, Imberger G, Walsh M, Chu R, Gluud C, Wetterslev J, Guyatt G, Devereaux PJ, Thabane L. The number of patients and events required to limit the risk of overestimation of intervention effects in meta-analysis--a simulation study. PLoS One 2011;6:e25491.
60. Turner RM, Bird SM, Higgins JP. The impact of study size on meta-analyses: examination of underpowered studies in Cochrane reviews. PLOS One 2013;8:e59202.
61. Usichenko TI, Röttenbacher I, Kohlmann T, Jülich A, Lange J, Mustea A, Engel G, Wendt M. Implementation of the quality management system improves postoperative pain treatment: a prospective pre-/post-interventional questionnaire study. Br J Anaesth 2013;110:87-95.
62. Visentin M. [Towards a pain-free hospital. A project and a campaign for the improvement of health care]. Recenti Prog Med 1999;90:321-4.
63. Vos T, Flaxman AD, Naghavi M, Lozano R, Michaud C, Ezzati M, Shibuya K, Salomon JA, Abdalla S, Aboyans V, Abraham J, Ackerman I, Aggarwal R, Ahn SY, Ali MK, Alvarado M, Anderson HR, Anderson LM, Andrews KG, Atkinson C, Baddour LM, Bahalim AN, Barker-Collo S, Barrero LH, Bartels DH, Basáñez MG, Baxter A, Bell ML, Benjamin EJ, Bennett D, Bernabé E, Bhalla K, Bhandari B, Bikbov B, Bin Abdulhak A, Birbeck G, Black JA, Blencowe H, Blore JD, Blyth F, Bolliger I, Bonaventure A, Boufous S, Bourne R, Boussinesq M, Braithwaite T, Brayne C, Bridgett L, Brooker S, Brooks P, Brugha TS, Bryan-Hancock C, Bucello C, Buchbinder R, Buckle G, Budke CM, Burch M, Burney P, Burstein R, Calabria B, Campbell B, Canter CE,

Carabin H, Carapetis J, Carmona L, Cella C, Charlson F, Chen H, Cheng AT, Chou D, Chugh SS, Coffeng LE, Colan SD, Colquhoun S, Colson KE, Condon J, Connor MD, Cooper LT, Corriere M, Cortinovis M, de Vaccaro KC, Couser W, Cowie BC, Criqui MH, Cross M, Dabhadkar KC, Dahiya M, Dahodwala N, Damsere-Derry J, Danaei G, Davis A, De Leo D, Degenhardt L, Dellavalle R, Delossantos A, Denenberg J, Derrett S, Des Jarlais DC, Dharmaratne SD, Dherani M, Diaz-Torne C, Dolk H, Dorsey ER, Driscoll T, Duber H, Ebel B, Edmond K, Elbaz A, Ali SE, Erskine H, Erwin PJ, Espindola P, Ewoigbokhan SE, Farzadfar F, Feigin V, Felson DT, Ferrari A, Ferri CP, Fèvre EM, Finucane MM, Flaxman S, Flood L, Foreman K, Forouzanfar MH, Fowkes FG, Franklin R, Fransen M, Freeman MK, Gabbe BJ, Gabriel SE, Gakidou E, Ganatra HA, Garcia B, Gaspari F, Gillum RF, Gmel G, Gosselin R, Grainger R, Groeger J, Guillemin F, Gunnell D, Gupta R, Haagsma J, Hagan H, Halasa YA, Hall W, Haring D, Haro JM, Harrison JE, Havmoeller R, Hay RJ, Higashi H, Hill C, Hoen B, Hoffman H, Hotez PJ, Hoy D, Huang JJ, Ibeanusi SE, Jacobsen KH, James SL, Jarvis D, Jasrasaria R, Jayaraman S, Johns N, Jonas JB, Karthikeyan G, Kassebaum N, Kawakami N, Keren A, Khoo JP, King CH, Knowlton LM, Kobusingye O, Koranteng A, Krishnamurthi R, Lalloo R, Laslett LL, Lathlean T, Leasher JL, Lee YY, Leigh J, Lim SS, Limb E, Lin JK, Lipnick M, Lipshultz SE, Liu W, Loane M, Ohno SL, Lyons R, Ma J, Mabweijano J, MacIntyre MF, Malekzadeh R, Mallinger L, Manivannan S, Marcenes W, March L, Margolis DJ, Marks GB, Marks R, Matsumori A, Matzopoulos R, Mayosi BM, McAnulty JH, McDermott MM, McGill N, McGrath J, Medina-Mora ME, Meltzer M, Mensah GA, Merriman TR, Meyer AC, Miglioli V, Miller M, Miller TR, Mitchell PB, Mocumbi AO, Moffitt TE, Mokdad AA, Monasta L, Montico M, Moradi-Lakeh M, Moran A, Morawska L, Mori R, Murdoch ME, Mwaniki MK, Naidoo K, Nair MN, Naldi L, Narayan KM, Nelson PK, Nelson RG, Nevitt MC, Newton CR, Nolte S, Norman

P, Norman R, O'Donnell M, O'Hanlon S, Olives C, Omer SB, Ortblad K, Osborne R, Ozgediz D, Page A, Pahari B, Pandian JD, Rivero AP, Patten SB, Pearce N, Padilla RP, Perez-Ruiz F, Perico N, Pesudovs K, Phillips D, Phillips MR, Pierce K, Pion S, Polanczyk GV, Polinder S, Pope CA 3rd, Popova S, Porrini E, Pourmalek F, Prince M, Pullan RL, Ramaiah KD, Ranganathan D, Razavi H, Regan M, Rehm JT, Rein DB, Remuzzi G, Richardson K, Rivara FP, Roberts T, Robinson C, De Leòn FR, Ronfani L, Room R, Rosenfeld LC, Rushton L, Sacco RL, Saha S, Sampson U, Sanchez-Riera L, Sanman E, Schwebel DC, Scott JG, Segui-Gomez M, Shahraz S, Shepard DS, Shin H, Shivakoti R, Singh D, Singh GM, Singh JA, Singleton J, Sleet DA, Sliwa K, Smith E, Smith JL, Stapelberg NJ, Steer A, Steiner T, Stolk WA, Stovner LJ, Sudfeld C, Syed S, Tamburlini G, Tavakkoli M, Taylor HR, Taylor JA, Taylor WJ, Thomas B, Thomson WM, Thurston GD, Tleyjeh IM, Tonelli M, Towbin JA, Truelsen T, Tsilimbaris MK, Ubeda C, Undurraga EA, van der Werf MJ, van Os J, Vavilala MS, Venketasubramanian N, Wang M, Wang W, Watt K, Weatherall DJ, Weinstock MA, Weintraub R, Weisskopf MG, Weissman MM, White RA, Whiteford H, Wiersma ST, Wilkinson JD, Williams HC, Williams SR, Witt E, Wolfe F, Woolf AD, Wulf S, Yeh PH, Zaidi AK, Zheng ZJ, Zonies D, Lopez AD, Murray CJ, AlMazroa MA, Memish ZA. Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet 2012;380(9859):2163-96.

64. Watt-Watson J, Stevens B, Garfinkel P, Streiner D, Gallop R. Relationship between nurses' pain knowledge and pain management outcomes for their postoperative cardiac patients. J Adv Nurs 2001;36:535-45.

## Tables

Table 1: Clinical setting and geographical distribution of studies, according to study size

Table 2: Definitions of discordance

Table 3: Direction of disagreement

## Supplementary files

Supplementary file 1: References to included studies

Supplementary file 2: Design and results of individual studies

Supplementary file 3: Number of studies and patients by decade and size of study